(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 109 132 A2

# (12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
20.06.2001 Bulletin 2001/25

(51) Int Cl.$^7$: G06T 7/60

(21) Application number: 00204309.9

(22) Date of filing: 04.12.2000

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 14.12.1999 US 460759

(71) Applicant: EASTMAN KODAK COMPANY
Rochester, New York 14650 (US)

(72) Inventors:
- Savakis, Andreas E., c/o Eastman Kodak Company Rochester, New York 14650-2201 (US)
- Etz, Stephen, c/o Eastman Kodak Company Rochester, New York 14650-2201 (US)

(74) Representative: Parent, Yves
KODAK INDUSTRIE
Département Brevets - CRT
Zone Industrielle
B.P. 21
71102 Chalon-sur-Saône Cédex (FR)

## (54) Method for automatic assessment of emphasis and appeal in consumer images

(57) An image is automatically assessed with respect to certain features, wherein the assessment is a determination of the degree of importance, interest or attractiveness of the image. First, a digital image is obtained corresponding to the image. Then one or more quantities are computed that are related to one or more features in the digital image, including one or more features pertaining to the content of the digital image. The quantities are processed with a reasoning algorithm that is trained on the opinions of one or more human observers, and an output is obtained from the reasoning algorithm that assesses the image. More specifically, the reasoning algorithm is a Bayesian network that provides a score which, when done for a group of images, selects one image as the emphasis image. The features pertaining to the content of the digital image include people-related features and/or subject-related features. Moreover, additional quantities may be computed that relate to objective measures of the digital image, such as colorfulness and/or sharpness.

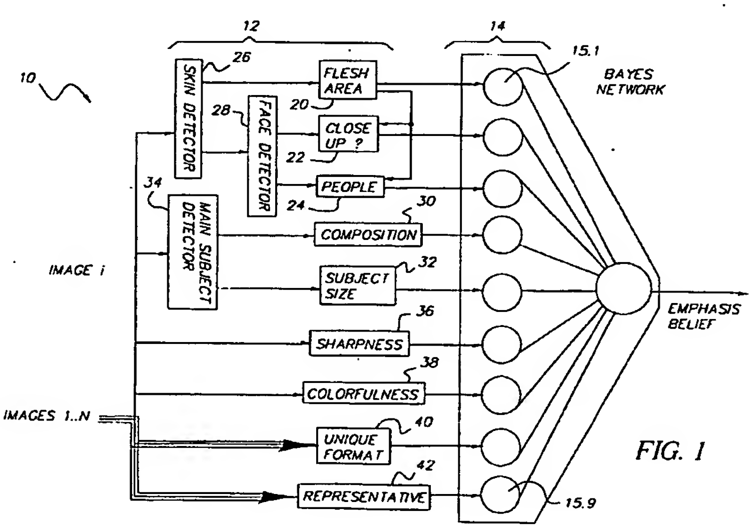


FIG. 1

EP 1 109 132 A2

Printed by Jouve, 75001 PARIS (FR)

**Description**

## FIELD OF THE INVENTION

**[0001]** The invention relates generally to the field of image processing, and in particular to the field of image assessment and understanding.

## BACKGROUND OF THE INVENTION

**[0002]** Image assessment and understanding deal with problems that are easily solved by human beings given their intellectual faculties but are extremely difficult to solve by fully automated computer systems. Image understanding problems that are considered important in photographic applications include main subject detection, scene classification , sky and grass detection, people detection, automatic detection of orientation, etc. In a variety of applications that deal with a group of pictures, it is important to rank the images in terms of a logical order, so that they can be processed or treated according to their order. A photographic application currently of interest is automatic albuming, where a group of digital images are automatically organized into digital photo albums. This involves clustering the images into separate events and then laying out each event in some logical order, if possible. This order implies at least some attention to the relative content of the images, i.e., based on the belief that some images would likely be preferred over others.

**[0003]** A number of known algorithms, such as dud detection, event detection and page layout algorithms, are useful in connection with automatic albuming applications. Dud detection addresses the elimination, or de-emphasis, of duplicate images and poor quality images, while event detection involves the clustering of images into separate events by certain defined criteria, such as date and time. Given a set of images that belong to the same event, the objective of page layout is to layout each event in some logical and pleasing presentation, e.g., to find the most pleasing and space-efficient presentation of the images on each page. It would be desirable to be able to select the most important image in the group of images, e.g., the one that should receive the most attention in a page layout.

**[0004]** Due to the nature of the image assessment problem, i.e., that an automated system is expected to generate results that are representative of high-level cognitive human (understanding) processes, the design of an assessment system is a challenging task. Effort has been devoted to evaluating text and graphical data for its psychological effect, with the aim of creating or editing a document for a particular visual impression (see, e.g., U.S. Patent Nos. 5,875,265 and 5,424,945). In the '265 patent, a system analyzes an image, in some case with the aid of an operator, to determine correspondence of visual features to sensitive language that is displayed for use by the operator. The difficulty in this system is that the visual features are primarily based on low level features, i.e., color and texture, that are not necessarily related to image content, and a language description is difficult is to use for relative ranking of images. The '945 patent discloses a system for evaluating the psychological effect of text and graphics in a document. The drawback with the '945 patent is that it evaluates the overall visual impression of the document, without regard to its specific content, which reduces its usefulness for developing relative ranking. Besides their complexity and orientation toward discernment of a psychological effect, these systems focus on the analysis and creation of a perceptual impression rather than on the assessment and utilization of an existing image.

## SUMMARY OF THE INVENTION

**[0005]** The present invention is directed to overcoming one or more of the problems set forth above. Briefly summarized, according to one aspect of the present invention, a method is disclosed for assessing an image with respect to certain features, wherein the assessment is a determination of the degree of importance, interest or attractiveness of an image. First, a digital image is obtained corresponding to the image. Then one or more quantities are computed that are related to one or more features in the digital image, including one or more features pertaining to the content of the digital image. The quantities are processed with a reasoning algorithm that is trained on the opinions of one or more human observers, and an output is obtained from the reasoning algorithm that assesses the image. In a dependent aspect of the invention, the features pertaining to the content of the digital image include at least one of people-related features and subject-related features. Moreover, additional quantities may be computed that relate to one or more objective measures of the digital image, such as colorfulness or sharpness.

**[0006]** From another aspect, the invention may be seen as assessing the emphasis or appeal of an image, as hereinafter defined. From this perspective, for both appeal and emphasis assessment, self-salient image features are calculated, such as:

a. People related features: the presence or absence of people, the amount of skin or face area and the extent of close-up based on face size.

b. Objective features: the colorfulness and sharpness of the image.
c. Subject related features: the size of main subject and the goodness of composition based on main subject mapping.

While the above-noted features are adequate for emphasis assessment, it is preferable that certain additional relative-salient image features are considered for appeal assessment, such as:

a. The representative value of each image in terms of color content.
b. The uniqueness of the picture aspect format of each image.

**[0007]** An assessment of an image is obtained with a reasoning engine, such as a Bayesian network, which accepts as input the above-noted features and is trained to generate image assessment values. This assessment may be an intrinsic assessment for individual images, in which case the self-salient features are processed by a Bayesian network trained to generate the image appeal values, or the assessment may be a relative assessment for a group of images, in which case the self-salient and, optionally, the relative-salient features are processed by a Bayesian network trained to generate image emphasis values.

**[0008]** The advantage of the invention lies in its ability to perform an assessment of one or more images without human intervention. In a variety of applications that deal with a group of pictures, such as automatic albuming, such an algorithmic assessment enables the automatic ranking of images in terms of their logical order, so that they can be more efficiently processed or treated according to their order.

**[0009]** These and other aspects, objects, features and advantages of the present invention will be more clearly understood and appreciated from a review of the following detailed description of the preferred embodiments and appended claims, and by reference to the accompanying drawings.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0010]** FIG. 1 is a block diagram of a network for calculating an emphasis value for an image.

**[0011]** FIG. 2 is a block diagram of a network for calculating an appeal value for an image.

**[0012]** FIG. 3 is a block diagram showing in more detail the components of main subject detection as shown in Figures 1 and 2.

**[0013]** FIG. 4 is a block diagram of a network architecture for calculating the relative emphasis values of a group of images.

**[0014]** FIGS. 5A- 5D are detailed diagrams of the component methods shown in Figure 3 for main subject detection.

**[0015]** FIG. 6 is a detailed diagram of a method for determining the colorfulness of an image.

**[0016]** FIG. 7 is a diagram of chromaticity plane wedges that are used for the colorfulness feature computation.

**[0017]** FIG. 8 is a block diagram of a method for skin and face detection.

**[0018]** FIG. 9 is a detailed block diagram of main subject detection as shown in Figure 5.

**[0019]** FIG. 10 is a diagram of a two level Bayesian net used in the networks shown in Figures 1 and 2.

**[0020]** FIG. 11 is a perspective diagram of a computer system for practicing the invention set forth in the preceding figures.

## DETAILED DESCRIPTION OF THE INVENTION

**[0021]** In the following description, a preferred embodiment of the present invention will be described as a software program. Those skilled in the art will readily recognize that the equivalent of such software may also be constructed in hardware. Because image processing algorithms and systems are well known, the present description will be directed in particular to algorithms and systems forming part of, or cooperating more directly with, the method in accordance with the present invention. Other aspects of such algorithms and systems, and hardware and/or software for producing and otherwise processing the image signals involved therewith, not specifically shown or described herein may be selected from such systems, algorithms, components and elements thereof known in the art. Given the description as set forth in the following specification, all software implementation thereof as a computer program is conventional and within the ordinary skill in such arts.

**[0022]** Still further, as used herein, the computer program may be stored in a computer readable storage medium, which may comprise, for example; magnetic storage media such as a magnetic disk (such as a floppy disk) or magnetic tape; optical storage media such as an optical disc, optical tape, or machine readable bar code; solid state electronic storage devices such as random access memory (RAM), or read only memory (ROM); or any other physical device or medium employed to store a computer program.

**[0023]** In a variety of applications that deal with a group of pictures, it is important to rank the images in terms of

their relative value and/or their intrinsic value, so that they can be processed or treated according to these values. As mentioned before, a photographic application currently of interest is automatic albuming, where a group of digital images are automatically organized into digital photo albums. This involves clustering the images into separate events and then laying out each event in some logical order. This logical order may be based upon two related assessments of the images: image appeal and image emphasis. Image appeal is the intrinsic degree of importance, interest or attractiveness of an individual picture; image emphasis, on the other hand, is the relative importance, interest or attractiveness of the picture with respect to other pictures in the event or group.

**[0024]** Once the assessments are obtained, it would be desirable to select the most important image in the group of images, e.g., the one that should receive the most attention in a page layout. Therefore an image assessment algorithm would fit well into the automatic albuming architecture between the steps of event detection and page layout. The event detection algorithm would output groups of images, where all images in a group belong to the same event. The assessment algorithm would be expected to operate on the images of each event and assign assessment values (i.e., emphasis and/or appeal values) to each image. The assessment values may be viewed as metadata that are associated with every image in a particular group and may be exploited by other algorithms. In such a proposed system, the page layout algorithm would take as input the relative assessment values of all images in each event, so that the images can be appropriately distributed between pages and arranged on each page. For example, emphasis values for a group of pictures in an event are used by a page layout algorithm to assign relative size and location of the pictures on album pages.

**[0025]** However, in any proposed system there are many open questions about the type of system architecture and the selection of effective features for evaluation. An architecture that has been successfully applied to other image understanding problems is based on a feature extraction stage followed by a classification stage. With respect to feature extraction, it is necessary to select an ensemble of features. For this there are two likely approaches. One is to select features that intuitively appear to have some relevance to image assessment values. The problem with this approach is that there is no good justification for the selection of features. A second approach is to base feature selection on experience gained through controlled experiments. Since the inventors found no such experiments on record, a ground truth study was conducted to obtain data that would point to meaningful features. The results of the ground truth study are used for feature selection and for training the classifier.

**[0026]** Referring first to Figure 1, an image emphasis network 10 for computing an emphasis value is shown to comprise two stages: a feature extraction stage 12 and a classification stage 14. The feature extraction stage 12 employs a number of algorithms, each designed to measure some image feature characteristic, where a quantitative measure of the feature is expressed by the value of the output of the algorithm. The outputs of the feature extraction stage 12 thus represent statistical evidence of the presence (or absence) of certain features; the outputs are then integrated by the classification stage 14 to compute an emphasis value. This value may, e.g., range from 0 to 100 and indicates the likelihood or belief that the processed image is the emphasis image. After the emphasis values have been computed for a group of images in separate image emphasis networks 10.1, 10.2...10.N, as shown in Figure 4, the emphasis values are compared in a comparator stage 16 and normalized in respective normalization stages 16.1, 16.2 ... 16.N. The image with the highest emphasis value is chosen as the emphasis image for the group.

**[0027]** An ensemble of features was selected for the feature extraction stage 12 on the basis of ground truth studies of the preference of human observers. The ground truth studies showed that the features that are important for emphasis image selection are not strongly related to traditional image quality metrics, such as sharpness, contrast, film grain and exposure, although one or more of these traditional metrics may continue to have value in the calculation of an assessment value. The selected features may be generally divided into three categories: (a) features related to people, (b) features related to the main subject, and (c) features related to objective measures of the image. Referring to Figure 1, features related to people are extracted by a skin area detector 20, a close-up detector 22 and a people detector 24. The input image i is typically processed through a skin detector 26 and a face detector 28 to generate intermediate values suitable for processing by the people-related feature detectors 20, 22 and 24. The features related to the main subject are extracted by a composition detector 30 and a subject size detector 32, based on input from a main subject detector 34. The composition detector 30 is composed of several composition-related main subject algorithms, as shown in Figure 3, including a main subject variance algorithm 30.1, a main subject centrality algorithm 30.2 and a main subject compactness algorithm 30.3. The main subject data is clustered in a clustering stage 31 and then provided to the composition-related algorithms 30.2 and 30.3 and to the subject size algorithm 32. The features related to objective measures of the image are extracted by a sharpness detector 36, a colorfulness detector 38 and a unique format detector 40. In addition, an objective measure related to how representative the color content of an image is relative to a group of images is extracted by a representative color detector 42.

**[0028]** The feature ensemble shown in Figure 1 is used to calculate a value representative of image emphasis, which is defined as the degree of relative importance, interest or attractiveness of an image with respect to other images in a group. Since each image must be evaluated in relation to other images in a group, the image emphasis calculation thus embodies a network of image emphasis networks 10.1, 10.2....10.N, such as shown in Figure 4, which scores the

images as to their respective emphasis values. In practice, there may be but one image emphasis network 10, which is repeatedly engaged to determine the image emphasis value of a series of images; in this case, the sequentially obtained results could be stored in an intermediate storage (not shown) for input to the comparator 16. The feature ensemble shown in Figure 2, which is a subset of the feature ensemble shown in Figure 1, is used to calculate a value representative of image appeal, which is defined as the intrinsic degree of importance, interest or attractiveness of an image in an absolute sense, that is, without reference to other images. The features shown in Figure 2 are thus referred to as self-salient features, inasmuch as these features can stand on their own as an assessment of an image. In comparison, two additional features are detected in Figure 1, namely, the unique format feature and the representative color feature; these are referred to as relative-salient features, inasmuch as these features are measurements that necessarily relate to other images. (These features, however, are optional insofar as a satisfactory measure of emphasis can be obtained from the self-salient features alone.) Consequently, an assessment of both appeal and emphasis involve self-salient features, while only an assessment of emphasis may involve relative-salient features.

**[0029]** The extraction of the feature ensembles according to Figures 1 and 2 involves the computation of corresponding feature quantities, as set forth below.

## Objective Features

**[0030]** Objective features are the easiest to compute and provide the most consistent results in comparison to other types of features. Methods for computing them have been available for some time, and a large art of imaging science is based on such measures. Although a large number of objective features could potentially be computed, only colorfulness and sharpness are considered for purposes of both image emphasis and appeal (Figures 1 and 2), and additionally unique format and representative color for purposes of image emphasis (Figure 1). Other objective measures, such as contrast and noise, may be found useful in certain situations and are intended to be included within the coverage of this invention.

## Colorfulness

**[0031]** The colorfulness detector 38 provides a quantitative measure of colorfulness based on the observation that colorful pictures have colors that display high saturation at various hues. This was determined in ground truth studies by examining for the presence of high saturation colors along various hues. The assumption of sRGB color space was made with respect to the image data. In particular, and as shown in Figure 6, the colorfulness detector 38 implements the following steps for computing colorfulness. Initially, in step 60, the input image values i are transformed to a luminance/chrominance space. While many such transformations are known to the skilled person and may be used with success in connection with the invention, the preferred transformation is performed according to the following expressions:

$$\text{Neutral} = \left(\frac{R+G+B}{3}\right)$$

$$\text{Green-Magenta} = \left(\frac{2G-R-B}{4}\right)$$

$$\text{Illumination} = \left(\frac{B-R}{2}\right)$$

where neutral is a measure of luminance, and green-magenta and illumination are a measure of chrominance. In step 62, the chrominance plane (illumination, green-magenta) is divided and quantized into twelve chromaticity plane wedges, as shown in Figure 7, which are referred to as angular bins. Next, in step 64, each pixel is associated with one of the angular bins if its chrominance component lies within the bounds of that bin. The level of saturation (which is the distance from origin) is calculated in step 66 for each pixel in each angular bin. The number of high saturation pixels that populate each angular bin are then measured in step 68, where a high saturation pixel is one whose distance from the origin in the chrominance plane is above a certain threshold $T_c$ (e.g., $T_c$=0.33). For each angular bin, the bin is determined to be active in step 70 if the number of high saturation pixels exceeds a certain threshold $T_c$ (e.g., $T_c$=250 pixels). Colorfulness is then calculated in step 72 according to the following expression:

$$\text{Colorfulness} = \min\left\{\frac{\text{Number of active bins}}{10}\ ,\ 1.0\right\}$$

Note that this definition of colorfulness assumes that if 10 out of the 12 bins are populated, colorfulness is considered to be 1.0 and the image is most colorful.

Sharpness

**[0032]** The sharpness detector 36 implements the following steps to find sharpness features in the image:

a) The image is cropped at a 20% level along the border and converted to grayscale by extracting the green channel;
b) The image edges are detected in the green channel using a Sobel operator after running a 3x3 averaging filter to reduce noise;
c) An edge histogram is formed and the regions that contain the strongest edges are identified as those that are above the 90$^{th}$ percentile of the edge histogram;
d) The strongest-edge regions are refined through median filtering, and the statistics of the strongest edges are computed; and
e) The average of the strongest edges provides an estimate of sharpness.

Further details of the method employed for sharpness detection may be found in commonly assigned U.S. Serial No. 09/274,645, entitled "A Method for Automatically Detecting Digital Images that are Undesirable for Placing in Albums", filed March 23, 1999 in the names of Andreas Savakis and Alexander Loui, and which is incorporated herein by reference.

Format Uniqueness

**[0033]** Participants in the ground truth experiment indicated that pictures taken in APS "panoramic" mode are more deserving of emphasis. Preliminary analysis of the ground truth data indicated that if a picture was the *only* panoramic picture in a group, this fact increases its likelihood of being selected as the emphasis image. The relative feature "format uniqueness" represents this property.

**[0034]** The unique format detector 40 implements the following algorithm for each image $i$ in the group, in which the format f is based on the long and short pixel dimensions $l_i$, $s_i$, of the image:

$$f_i \equiv \begin{cases} \mathbf{C}, & l_i / s_i < 1.625, \\ \mathbf{H}, & 1.625 \le l_i / s_i < 2.25, \\ \mathbf{P}, & 2.25 \le l_i / s_i. \end{cases}$$

Then format uniqueness U is

$$Ui = \begin{cases} 1, & f_i \ne f_j, \forall i \ne j, \\ 0, & \text{otherwise.} \end{cases}$$

Representative Color

**[0035]** The representative color detector 42 implements the following steps to determine how representative the color of an image is:

1. For each image $i$, compute the color histogram $h_i(R, G, B)$ (in RGB or Luminance/Chrominance space)
2. Find the average color histogram for the group by averaging all of the image histograms as follows:

$$A_h(R,G,B) = \sum_{i=1}^{N} h_i(R,G,B)$$

3. For each image $i$, compute the distance between the histogram of the image and the average color histogram (Euclidian or Histogram intersection distance), as follows:

$$d_i(R,G,B) = \frac{1}{2}\sum_{i=1}^{N} |h_i(R,G,B) - A_h(R,G,B)|$$

4. Find the maximum of the distances computed in 3, as follows:.

$$d_{\max}(R,G,B) = \max_{i=1...N}\{d_i(R,G,B)\}$$

5. The representative measure $r$ is obtained by dividing each of the distances with the maximum distance (can vary from 0 to 1), as follows:

$$r_i(R,G,B) = \frac{d_i(R,G,B)}{d_{max}(R,G,B)}$$

People-Related Features

**[0036]** People related features are important in determining image emphasis, but many of the positive attributes that are related to people are difficult to compute, e.g. people smiling, people facing camera, etc. Skin detection methods allow the computation of some people-related features such as: whether people are present, the magnitude of the skin area, and the amount of closeup.

Skin and Face Detection

**[0037]** The skin detection method that is used by the skin detector 26, and the face detection method that is used by the face detector 28, is based on the method disclosed in commonly assigned patent application Serial No. 09/112,661 entitled "A Method for Detecting Human Faces in Digitized Images" which was filed July 9, 1998 in the names of H.C. Lee and H. Nicponski., and which is incorporated herein by reference.

**[0038]** Referring to Fig. 8, an overview is shown of the method disclosed in Serial No. 09/112,661. The input images are color balanced to compensate for predominant global illumination in step S102, which involves conversion from *(r, g,b)* values to *(L,s,t)* values. In the *(L,s,t)* space, the *L* axis represents the brightness of a color, while the *s* and *t* axes are chromatic axes. The *s* component approximately represents the illuminant variations from daylight to tungsten light, from blue to red. The *t* component represents an axis between green and magenta. A number of well-known color balancing algorithms may be used for this step, including a simple method of averaging-to-gray. Next, a k-mode clustering algorithm is used for color segmentation in step S104. A disclosure of this algorithm is contained in commonly assigned U.S. Patent 5,418,895, which is incorporated herein by reference. Basically, a 3-D color histogram in *(L,s,t)* space is formed from the input color image and processed by the clustering algorithm. The result of this step is a region map with each connected region having a unique label. For each region, the averaged luminance and chromaticity are computed in step S106. These features are used to predict possible skin regions (candidate skin regions) based on conditional probability and adaptive thresholding. Estimates of the scale and in-plane rotational pose of each skin region are then made by fitting a best ellipse to each skin region in step S108. Using a range of scales and in-plane rotational pose around these estimates, a series of linear filtering steps are applied to each facial region in step S110 for identifying tentative facial features . A number of probability metrics are used in step S112 to predict the likelihood that the region actually represents a facial feature and the type of feature it represents.

**[0039]** Features that pass the previous screening step are used as initial features in a step S114 for a proposed face. Using projective geometry, the identification of the three initial features defines the possible range of poses of the head. Each possible potential face pose, in conjunction with a generic threedimensional head model and ranges of variation of the position of the facial features, can be used to predict the location of the remaining facial features. The list of

candidate facial features can then be searched to see if the predicted features were located. The proximity of a candidate feature to its predicted location and orientation affects the probabilistic estimate of the validity of that feature.

**[0040]** A Bayesian network probabilistic model of the head is used in a step S116 to interpret the accumulated evidence of the presence of a face. The prior probabilities of the network are extracted from a large set of training images with heads in various orientations and scales. The network is initiated with the proposed features of the candidate face, with their estimated probabilities based on computed metrics and spatial conformity to the template. The network is then executed with these initial conditions until it converges to a global estimate of the probability of face presence. This probability can be compared against a hard threshold or left in probabilistic form when a binary assessment is not needed. Further details of this skin and face detection method may be found in Serial No. 09/112,661, which is incorporated herein by reference.

Skin Area

**[0041]** The percentage of skin/face area in a picture is computed by the skin area detector 20 on its own merit, and also as a preliminary step to people detection and close-up detection. Consequently, the output of the skin area detector 20 is connected to the classification stage 14 and also input to the close-up detector 22 and the people detector 24. Skin area is a continuous variable between 0 and 1 and correlates to a number of features related to people. For example, for pictures taken from the same distance, increasing skin area indicates that there are more people in the picture and correlates with the positive indicator of "whole group in photo." Alternatively, if two pictures contain the same number of people, larger skin area may indicate larger magnification, which correlates with the positive attribute of "closeup." Other explanations for larger skin area are also possible due to subject positioning.

Close-up

**[0042]** The close-up detector 22 employs the following measure for determining close-up:

a) skin detection is performed and the resulting map is examined at the central region (25% from border); and
b) close-up is determined as the percentage of skin area at the central portion of the image.

In some cases, face detection would be more appropriate than skin detection for determining close-up.

People Present

**[0043]** The presence of people is detected by the people detector 24 when a significant amount of skin area is present in the image. The percentage of skin pixels in the image is computed and people are assumed present when the skin percentage is above a threshold $T_f$ number of pixels (e.g., $T_f$ = 20 pixels). People present is a binary feature indicating the presence or absence of people for 1 or 0 respectively.

Composition features

**[0044]** Good composition is a very important positive attribute of picture emphasis and bad composition is the most commonly mentioned negative attribute. Automatic evaluation of the composition of an image is very difficult and sometimes subjective. Good composition may follow a number of general well-known rules, such as the rule of thirds, but these rules are often violated to express the photographer's perspective.

Main Subject Detection

**[0045]** The algorithm used by the main subject detector 34 is disclosed in commonly assigned patent application Serial No. 09/223,860 entitled "Method for Automatic Determination of Main Subjects in Consumer Images", filed December 31, 1998 in the names of J. Luo, S. Etz and A. Singhal. Referring to Fig. 9, there is shown a block diagram of an overview of the main subject detection method disclosed in Serial No. 09/223,860. First, an input image of a natural scene is acquired and stored in step S200 in a digital form. Then, the image is segmented in step S202 into a few regions of homogeneous properties. Next, the region segments are grouped into larger regions in step S204 based on similarity measures through non-purposive perceptual grouping, and further grouped in step S206 into larger regions corresponding to perceptually coherent objects through purposive grouping (purposive grouping concerns specific objects). The regions are evaluated in step S208 for their saliency using two independent yet complementary types of saliency features - structural saliency features and semantic saliency features. The structural saliency features, including a set of low-level early vision features and a set of geometric features, are extracted in step S208a, which are

further processed to generate a set of self-saliency features and a set of relative saliency features. Semantic saliency features in the forms of key subject matters, which are likely to be part of either foreground (for example, people) or background (for example, sky, grass), are detected in step S208b to provide semantic cues as well as scene context cues. The evidences of both types are integrated in step S210 using a reasoning engine based on a Bayes net to yield the final belief map step S212 of the main subject.

**[0046]** To the end of semantic interpretation of images, a single criterion is clearly insufficient. The human brain, furnished with its *a priori* knowledge and enormous memory of real world subjects and scenarios, combines different subjective criteria in order to give an assessment of the interesting or primary subject(s) in a scene. The following extensive list of features are believed to have influences on the human brain in performing such a somewhat intangible task as main subject detection: location, size, brightness, colorfulness, texturefulness, key subject matter, shape, symmetry, spatial relationship (surroundedness/occlusion), borderness, indoor/outdoor, orientation, depth (when applicable), and motion (when applicable for video sequence).

**[0047]** The low-level early vision features include color, brightness, and texture. The geometric features include location (centrality), spatial relationship (borderness, adjacency, surroundedness, and occlusion), size, shape, and symmetry. The semantic features include skin, face, sky, grass, and other green vegetation. Those skilled in the art can define more features without departing from the scope of the present invention. More details of the main subject detection algorithm are provided in Serial No. 09/223,860, which is incorporated herein by reference.

**[0048]** The aforementioned version of the main subject detection algorithm is computationally intensive and alternative versions may be used that base subject detection on a smaller set of subject-related features. Since all of the composition measures considered here are with respect to the main subject belief map, it is feasible to concentrate the system on the most computationally effective aspects of these measures, such as aspects bearing mostly on the "centrality" measure. These aspects are considered in judging the main subject, thereby reducing the overall computational complexity at the expense of some accuracy. It is a useful property of the Bayesian Network used in the main subject detection algorithm that features can be excluded in this way without requiring the algorithm to be retrained. Secondly, it takes advantage of the fact that images supplied to main subject detector 50 are known to be oriented right-side-up. The subject-related features associated with spatial location of a region within the scene can be modified to reflect this knowledge. For example, without knowing scene orientation the main subject detector 50 assumes a center-weighted distribution of main subject regions, but with known orientation a bottom-centerweighted distribution may be assumed.

**[0049]** Referring to Figure 3, after the main subject belief map has been computed in the main subject detector 50, it is segmented in a clustering stage 31 into three regions using k-means clustering of the intensity values. The three regions correspond to pixels that have high probability of being part of the main subject, pixels that have low probability of being part of the main subject, and intermediate pixels. Based on the quantized map, the features of main subject size, centrality, compactness, and interest (variance) are computed as described below in reference to Figs. 5A- SD.

Main Subject Variance

**[0050]** One way to characterize the contents of a photograph is by how interesting it is. For the purpose of emphasis image selection, an image with the following characteristics might be considered interesting.

- the main subject is interesting in and of itself, by virtue of its placement in the frame.
- the main subject constitutes a reasonably large area of the picture, but not the entire frame.
- the background does not include isolated objects that can distract from the main subject.

An estimate of the interest level of each image is computed by estimating the variance in the main subject map. This feature is primarily valuable as a counterindicator: that is, uninteresting images should not be the emphasis image. In particular, and as shown in Figure 5A, the main subject variance detector 30.1 implements the following steps for computing main subject variance. Initially, in step S10, the statistical variance v of all main subject belief map values is computed. In step S12, the main subject variance feature y is computed according to the formula:

$$y = \min(1, 2.5 * \mathrm{sqrt}(v)/127.5)$$

Main subject centrality

**[0051]** The main subject centrality is computed as the distance between the image center and the centroid of the high probability (and optionally the intermediate probability) region(s) in the quantized main subject belief map. In particular, and as shown in Figure 5B, the main subject centrality detector 30.2 implements the following steps for

computing main subject centrality. Initially, in step S20, the pixel coordinates of the centroid of the highest-valued cluster is located. In step S22, the Euclidean distance j from the center of the image to the centroid is computed. In step S24, the normalized distance k is computed by dividing j by the number of pixels along the shortest side of the image. In step S26, the main subject centrality feature m is computed according to the formula:

$$m = \min(k,1)$$

Main subject size

**[0052]** The size of the main subject is determined by the size of the high probability (and optionally the intermediate probability) region(s) in the quantized main subject belief map. It is expressed as the percentage of the central area (25% from border) that is occupied by the high (and optionally the intermediate) probability region. In particular, and as shown in Figure 5C, the main subject size detector 32 implements the following steps for computing main subject size. Initially, in step S30, the number of pixels fin the intersection of the highest-valued cluster and the rectangular central ¼ of the image area is counted. In step S32, the main subject size feature g is computed according to the formula:

$$g=f/N$$

where N is the total number of image pixels.

Main Subject Compactness

**[0053]** The compactness of the main subject is estimated by computing a bounding rectangle for the high probability (and optionally the intermediate probability) region(s) in the quantized main subject belief map, and then examining the percentage of the bounding rectangle that is occupied by the main subject. In particular, and as shown in Figure 5D, the main subject compactness detector 30.3 implements the following steps for computing main subject compactness. Initially, in step S40, the number of pixels a in the highest-valued cluster is counted. In step S42, the smallest rectangular box which contains all pixels in the highest-valued cluster (the bounding box) is computed, and in step S44 the area b of the bounding box, in pixels, is calculated. In step S46, the main subject compactness feature e is determined according to the formula:

$$e = \min(1, \max(0, 2*(a/b-0.2)))$$

where e will be a value between 0 and 1, inclusive.

Classification Stage

**[0054]** The feature quantities generated according to the algorithms set forth above are applied to the classification stage 14, which is preferably a reasoning engine that accepts as input the self-salient and/or the relative-salient features and is trained to generate image assessment (emphasis and appeal) values. Different evidences may compete or reinforce each according to knowledge derived from the results of the ground truth study of human observers--evaluations of real images. Competition and reinforcement are resolved by the inference network of the reasoning engine. A preferred reasoning engine is a Bayes network.

**[0055]** A Bayes net (see, e.g., J. Pearl, *Probabilistic Reasoning in Intelligent Systems,* San Francisco, CA: Morgan Kaufmann, 1988) is a directed acyclic graph that represents causality relationships between various entities in the graph, where the direction of links represents causality relationships between various entities in the graph, and where the direction of links represents causality. Evaluation is based on knowledge of the Joint Probability Distribution Function (PDF) among various entities. The Bayes net advantages include explicit uncertainty characterization, efficient computation, easy construction and maintenance, quick training, and fast adaptation to changes in the network structure and its parameters. A Bayes net consists of four components:

- Priors: The initial beliefs about various nodes in the Bayes net.
- Conditional Probability Matrices (CPMs): Expert knowledge about the relationship between two connected nodes in the Bayes net.

- Evidences: Observations from feature detectors that are input to the Bayes net.
- Posteriors: The final computed beliefs after the evidences have been propagated through the Bayes net.

**[0056]** The most important component for training is the set of CPMs, shown as CPM stages 15.1 ... 15.9 in Figure 1 (and 15.1 ... 15.7 in Figure 2) because they represent domain knowledge for the particular application at hand. While the derivation of CPMs is familiar to a person skilled in using reasoning engines such as a Bayes net, the derivation of an exemplary CPM will be considered later in this description.

**[0057]** Referring to Figures 1 and 2, a simple two-level Bayes net is used in the current system, where the emphasis (or appeal) score is determined at the root node and all the feature detectors are at the leaf nodes. It should be noted that each link is assumed to be conditionally independent of other links at the same level, which results in convenient training of the entire net by training each link separately, i.e., deriving the CPM for a given link independent of others. This assumption is often violated in practice; however, the independence simplification makes implementation feasible and produces reasonable results. It also provides a baseline for comparison with other classifiers or reasoning engines.

Probabilistic Reasoning

**[0058]** All the features are integrated by a Bayes net to yield the emphasis or appeal value. On one hand, different evidences may compete with or contradict each other. On the other hand, different evidences may mutually reinforce each other according to prior models or knowledge of typical photographic scenes. Both competition and reinforcement are resolved by the Bayes net-based inference engine.

**[0059]** Referring to Fig. 10, a two-level Bayesian net is used in the present invention that assumes conditional independence between various feature detectors. The emphasis or appeal value is determined at the root node 44 and all the feature detectors are at the leaf nodes 46. There is one Bayes net active for each image. It is to be understood that the present invention can be used with a Bayes net that has more than two levels without departing from the scope of the present invention.

Training Bayes nets

**[0060]** One advantage of Bayes nets is each link is assumed to be independent of links at the same level. Therefore, it is convenient for training the entire net by training each link separately, i.e., deriving the CPM 15.1... 15.9 for a given link independent of others. In general, two methods are used for obtaining CPM for each root-feature node pair:

1. Using Expert Knowledge

**[0061]** This is an ad-hoc method. An expert is consulted to obtain the conditional probabilities of each feature detector producing a high or low output given a highly appealing image.

2. Using Contingency Tables

**[0062]** This is a sampling and correlation method. Multiple observations of each feature detector are recorded along with information about the emphasis or appeal. These observations are then compiled together to create contingency tables which, when normalized, can then be used as the CPM 15.1... 15.9. This method is similar to neural network type of training (learning). This method is preferred in the present invention.

**[0063]** Consider the CPM for an arbitrary feature as an example. This matrix was generated using contingency tables derived from the ground truth and the feature detector. Since the feature detector in general does not supply a binary decision (referring to Table 1), fractional frequency count is used in deriving the CPM. The entries in the CPM are determined by

$$CPM = \left[ \left( \sum_{i \in I} \sum_{r \in R_i} n_i F_r^T T_r \right) P \right]^T \qquad (14)$$

$$F_r = \left[f^r_0 f^r_1 \cdots f^r_M\right], \quad T_r = \left[t^r_0 t^r_1 \cdots t^r_L\right],$$

$$P = diag\{p_j\}, \qquad p_j = \left(\sum_{i \in I} \sum_{r \in R_i} n_i t_r\right),$$

where I is the set of all training image groups, $R_i$ is the set of all images in group i, $n_i$ is the number of observations (observers) for group i. Moreover, $F_r$ represents an M-label feature vector for image r, $T_r$ represents an L-level ground-truth vector, and P denotes an L x L diagonal matrix of normalization constant factors. For example, in Table 1, images 1, 4, 5 and 7 contribute to boxes 00, 11, 10 and 01 in Table 2, respectively. Note that all the belief values have been normalized by the proper belief sensors. As an intuitive interpretation of the first column of the CPM for centrality, an image with a high feature value is about twice as likely to be highly appealing than not.

Table 1:

| An example of training the CPM. | | | |
|---|---|---|---|
| **Image Number** | **Ground Truth** | **Feature Detector Output** | **Contribution** |
| 1 | 0 | 0.017 | 00 |
| 2 | 0 | 0.211 | 00 |
| 3 | 0 | 0.011 | 00 |
| 4 | 0.933 | 0.953 | 11 |
| 5 | 0 | 0.673 | 10 |
| 6 | 1 | 0.891 | 11 |
| 7 | 0.93 | 0.072 | 01 |
| 8 | 1 | 0.091 | 01 |

Table 2:

| The trained CPM. | | |
|---|---|---|
| | Feature = 1 | feature = 0 |
| Emphasis or Appeal = 1 | 0.35 (11) | 0.65 (01) |
| Emphasis or Appeal = 0 | 0.17 (10) | 0.83 (00) |

**[0064]** While the invention has been described for use with a Bayes net, different reasoning engines may be employed in place of the Bayes net. For example, in Pattern Recognition and Neural Networks by B.D. Ripley (Cambridge University Press, 1996), a variety of different classifiers are described that can be used to solve pattern recognition problems, where having the right feature is normally the most important consideration. Such classifiers include linear discriminant analysis methods, flexible discriminants, (feed-forward) neural networks, non-parametric methods, tree-structured classifiers, and belief networks (such as Bayesian networks). It will be obvious to anyone of ordinary skill in such methods that any of these classifiers can be adopted as the reasoning engine for practice of the present invention.

Computer System

**[0065]** In describing the present invention, it should be apparent that the present invention is preferably utilized on any well-known computer system, such a personal computer. Consequently, the computer system will not be discussed in detail herein. It is also instructive to note that the images are either directly input into the computer system (for example by a digital camera) or digitized before input into the computer system (for example by scanning an original,

such as a silver halide film).

[0066] Referring to Fig. 11, there is illustrated a computer system 110 for implementing the present invention. Although the computer system 110 is shown for the purpose of illustrating a preferred embodiment, the present invention is not limited to the computer system 110 shown, but may be used on any electronic processing system. The computer system 110 includes a microprocessor-based unit 112 for receiving and processing software programs and for performing other processing functions. A display 114 is electrically connected to the microprocessor-based unit 112 for displaying user-related information associated with the software, e.g., by means of a graphical user interface. A keyboard 116 is also connected to the microprocessor based unit 112 for permitting a user to input information to the software. As an alternative to using the keyboard 116 for input, a mouse 118 may be used for moving a selector 120 on the display 114 and for selecting an item on which the selector 120 overlays, as is well known in the art.

[0067] A compact disk-read only memory (CD-ROM) 22 is connected to the microprocessor based unit 112 for receiving software programs and for providing a means of inputting the software programs and other information to the microprocessor based unit 112 via a compact disk 124, which typically includes a software program. In accordance with the invention, this software program could include the image assessment program described herein, as well as programs that utilize its output, such as the automatic albuming program. In addition, a floppy disk 126 may also include the software program, and is inserted into the microprocessor-based unit 112 for inputting the software program. Still further, the microprocessor-based unit 112 may be programmed, as is well known in the art, for storing the software program internally. The microprocessor-based unit 112 may also have a network connection 127, such as a telephone line, to an external network, such as a local area network or the Internet. The program could thus stored on a remote server and accessed therefrom, or downloaded as needed. A printer 128 is connected to the microprocessor-based unit 12 for printing a hardcopy of the output of the computer system 110.

[0068] Images may also be displayed on the display 114 via a personal computer card (PC card) 130, such as, as it was formerly known, a PCMCIA card (based on the specifications of the Personal Computer Memory Card International Association) which contains digitized images electronically embodied in the card 130. The PC card 130 is ultimately inserted into the microprocessor based unit 112 for permitting visual. display of the image on the display 114. Images may also be input via the compact disk 124, the floppy disk 126, or the network connection 127. Any images stored in the PC card 130, the floppy disk 126 or the compact disk 124, or input through the network connection 127, may have been obtained from a variety of sources, such as a digital camera (not shown) or a scanner (not shown).

[0069] The subject matter of the present invention relates to digital image understanding technology, which is understood to mean technology that digitally processes a digital image to recognize and thereby assign useful meaning to human understandable objects, attributes or conditions and then to utilize the results obtained in the further processing of the digital image.

[0070] The invention has been described with reference to a preferred embodiment. However, it will be appreciated that variations and modifications can be effected by a person of ordinary skill in the art without departing from the scope of the invention. While the invention has been described from time to time in connection with automatic albuming, it should be clear that there are many other uses for the invention, in particular any kind of image processing applications where an image needs to be evaluated for some process based on its relative or intrinsic value.

## Claims

1. A method for assessing an image with respect to certain features, wherein said assessment is a determination of the degree of importance, interest or attractiveness of an image, said method comprising the steps of:

   (a) obtaining a digital image corresponding to the image;
   (b) computing one or more quantities related to one or more features in the digital image, including one or more features pertaining to the content of the digital image;
   (c) processing said one or more quantities with a reasoning algorithm that is trained on the opinions of one or more human observers; and
   (d) providing an output from the reasoning algorithm that assesses the image.

2. The method as claimed in claim 1 wherein the features pertaining to the content of the digital image include at least one of people-related features and subject-related features.

3. The method as claimed in claim 1 wherein step (b) further includes computing one or more quantities related to one or more objective features pertaining to objective measures of the digital image.

4. The method as claimed in claim 3 wherein the objective features include at least one of colorfulness and sharpness.

5. The method as claimed in claim 1 wherein the reasoning algorithm in step (c) is trained at least in part from ground truth studies of candidate images.

6. The method as claimed in claim 1 wherein the reasoning algorithm is a Bayesian network.

7. The method as claimed in claim 1 wherein step (d) provides an output from the reasoning algorithm that scores the assessment of the image.

8. The method as claimed in claim 1 wherein step (a) obtains a plurality of digital images corresponding to a plurality of images and steps (b)-(d) are performed on each of the digital images such that an assessment is performed for each of the images.

9. The method as claimed in claim 8 wherein the assessments are ranked such that one image with the highest rank is selected as an emphasis image.

10. The method as claimed in claim 1 wherein the assessment of the image is used in an automatic albuming algorithm to arrange a page layout of images.

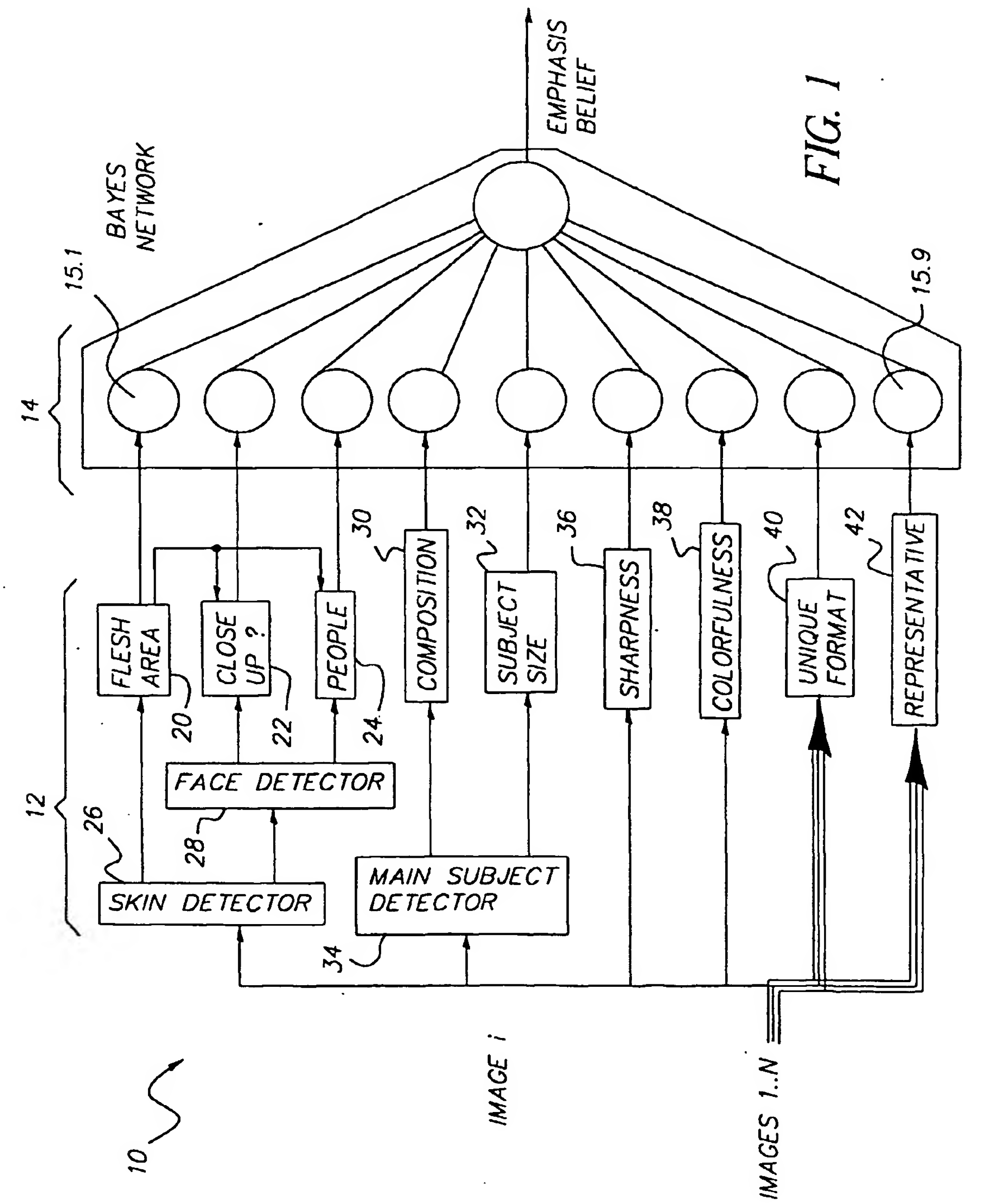
BAYES NETWORK
EMPHASIS BELIEF
15.1
15.9
14
12
10
IMAGE i
IMAGES 1..N
SKIN DETECTOR
26
FACE DETECTOR
28
MAIN SUBJECT DETECTOR
34
FLESH AREA
20
CLOSE UP ?
22
PEOPLE
24
COMPOSITION
30
SUBJECT SIZE
32
SHARPNESS
36
COLORFULNESS
38
UNIQUE FORMAT
40
REPRESENTATIVE
42

FIG. 1

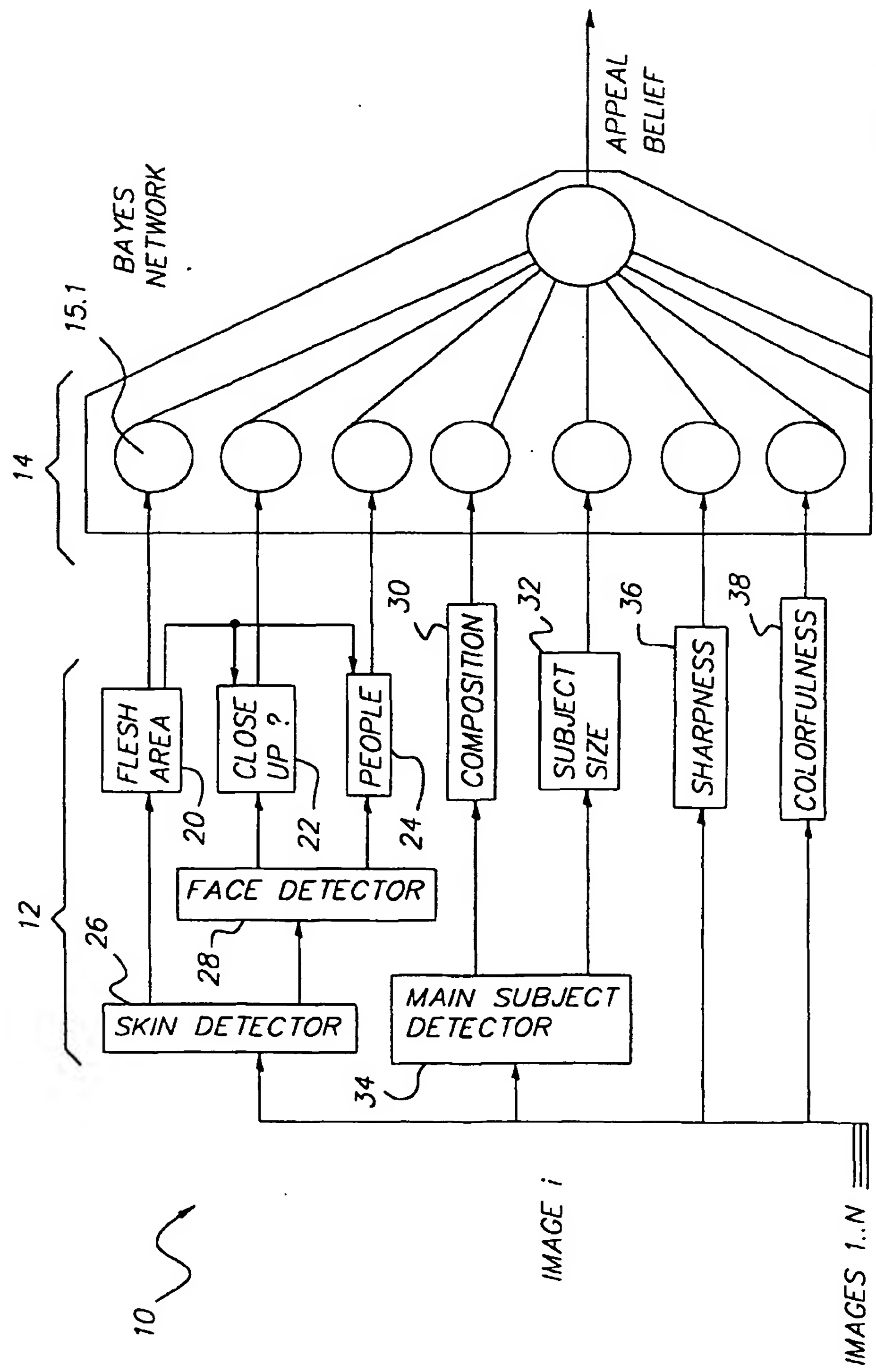
10
IMAGE i
IMAGES 1..N
SKIN DETECTOR
26
FACE DETECTOR
28
MAIN SUBJECT DETECTOR
34
FLESH AREA
20
CLOSE UP ?
22
PEOPLE
24
COMPOSITION
30
SUBJECT SIZE
32
SHARPNESS
36
COLORFULNESS
38
12
14
15.1
BAYES NETWORK
APPEAL BELIEF

FIG. 2

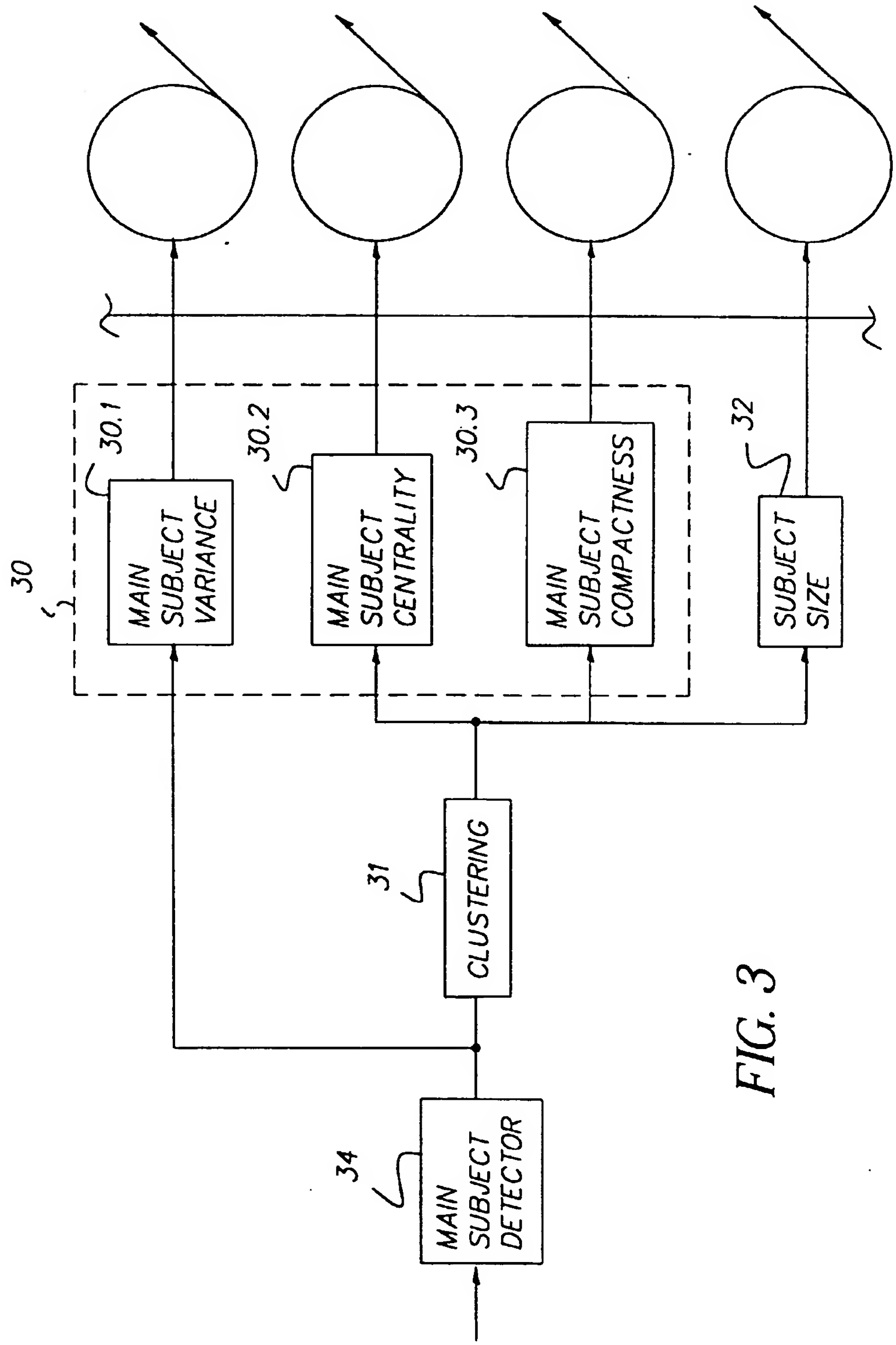
MAIN SUBJECT DETECTOR
34
CLUSTERING
31
MAIN SUBJECT VARIANCE
30.1
MAIN SUBJECT CENTRALITY
30.2
MAIN SUBJECT COMPACTNESS
30.3
SUBJECT SIZE
32
30

FIG. 3

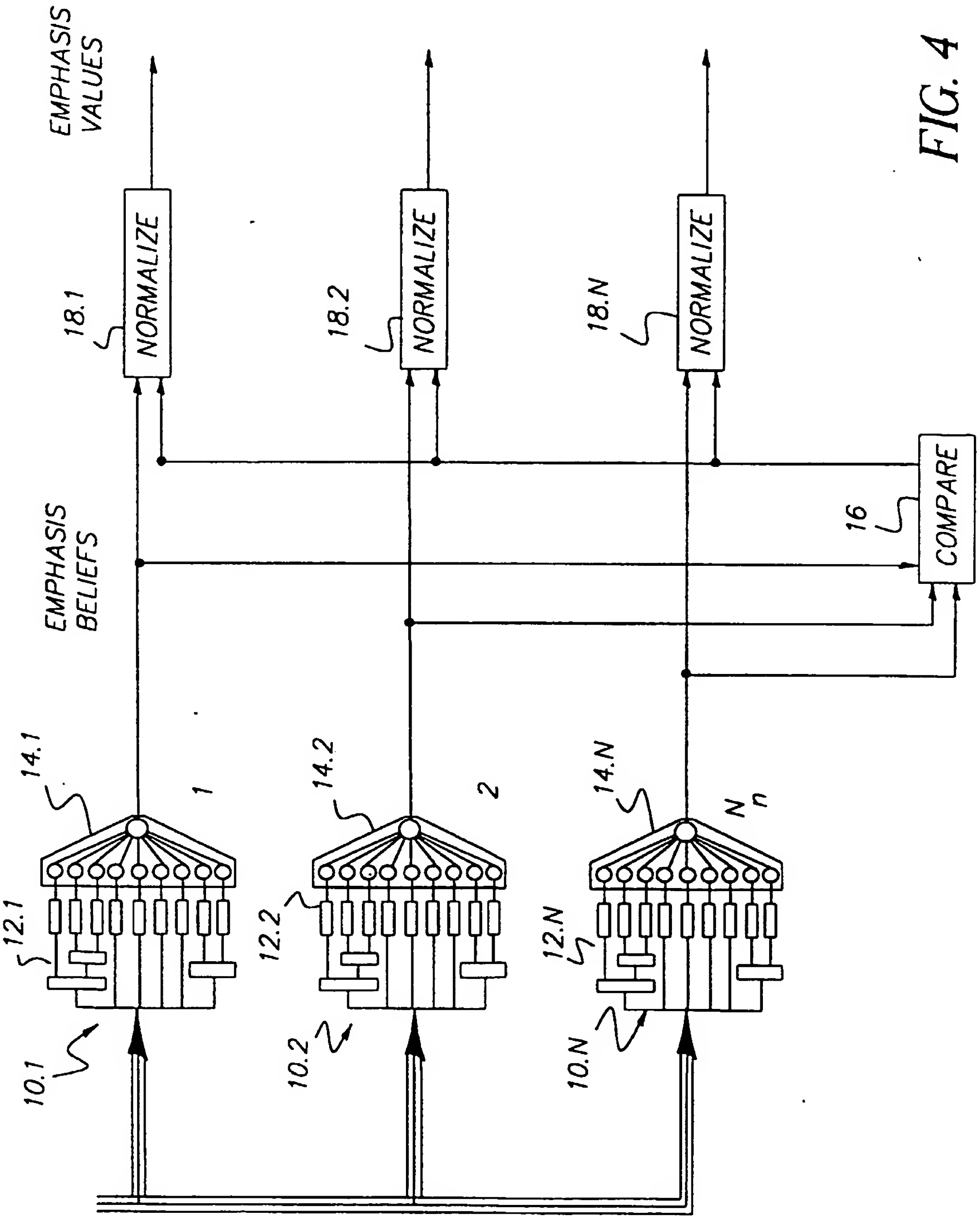
EMPHASIS VALUES
EMPHASIS BELIEFS
18.1
NORMALIZE
18.2
NORMALIZE
18.N
NORMALIZE
16
COMPARE
14.1
14.2
14.N
12.1
12.2
12.N
10.1
10.2
10.N
1
2
N
n

FIG. 4

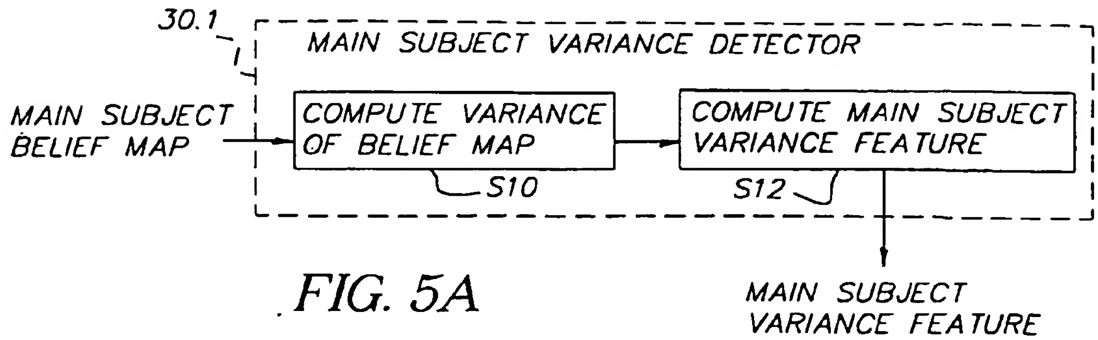


FIG. 5A

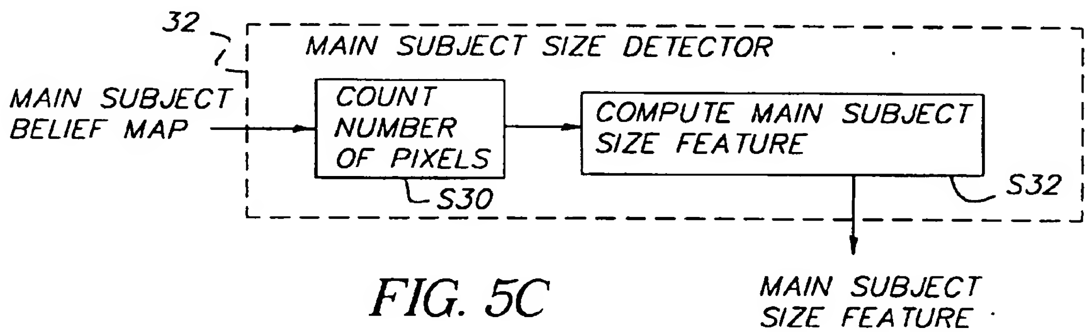


FIG. 5C

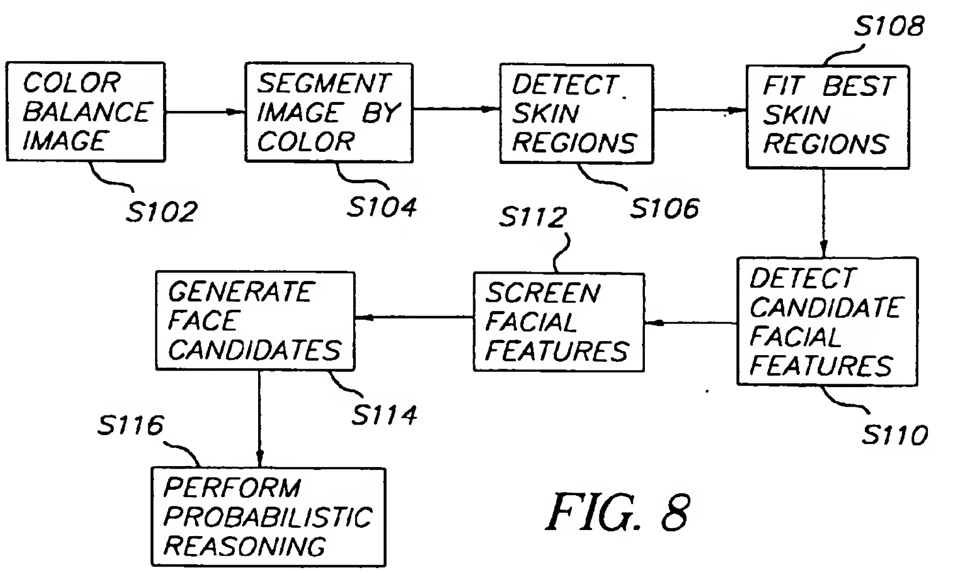


FIG. 8

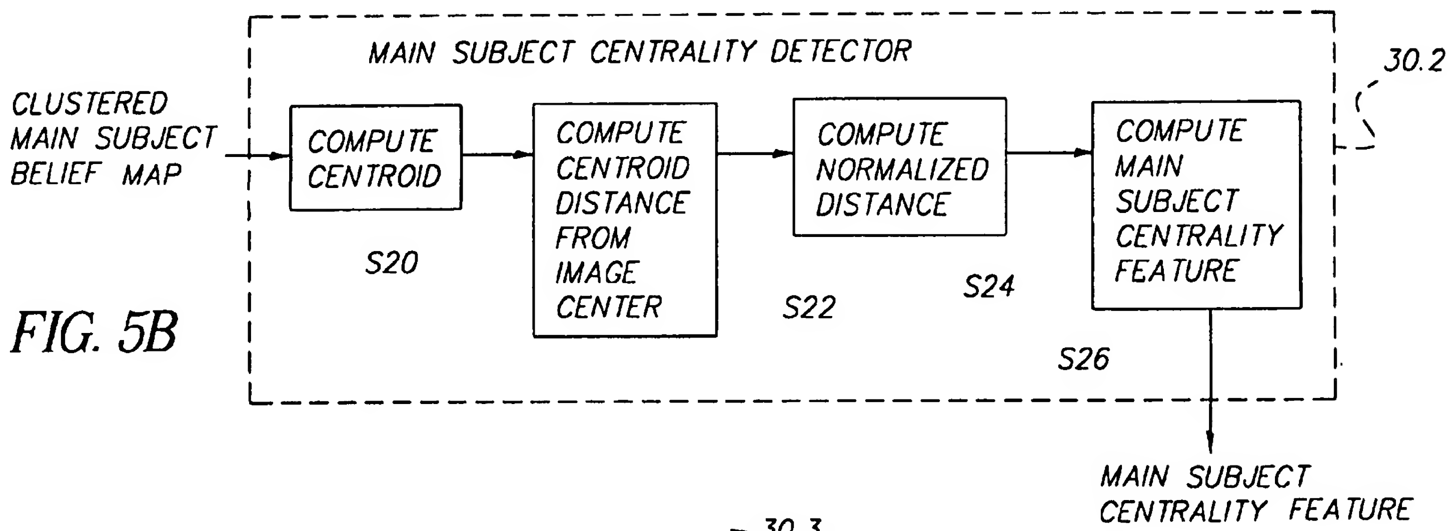
MAIN SUBJECT CENTRALITY DETECTOR
30.2
CLUSTERED MAIN SUBJECT BELIEF MAP
COMPUTE CENTROID
S20
COMPUTE CENTROID DISTANCE FROM IMAGE CENTER
S22
COMPUTE NORMALIZED DISTANCE
S24
COMPUTE MAIN SUBJECT CENTRALITY FEATURE
S26
MAIN SUBJECT CENTRALITY FEATURE

FIG. 5B

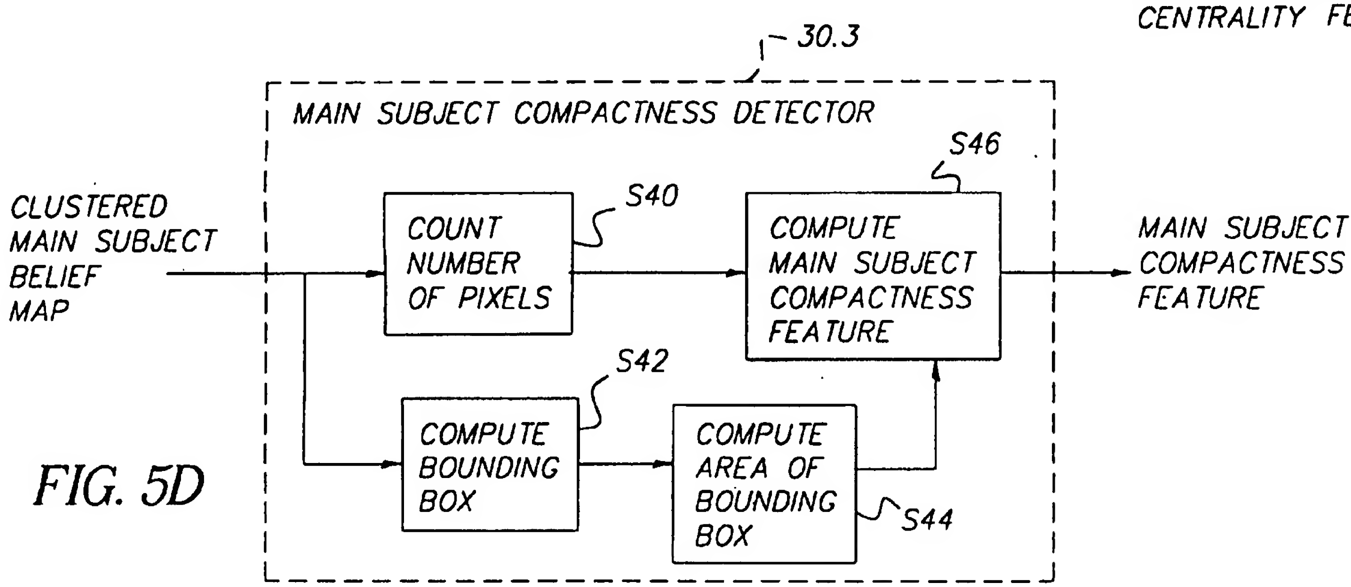
30.3
MAIN SUBJECT COMPACTNESS DETECTOR
CLUSTERED MAIN SUBJECT BELIEF MAP
COUNT NUMBER OF PIXELS
S40
COMPUTE MAIN SUBJECT COMPACTNESS FEATURE
S46
COMPUTE BOUNDING BOX
S42
COMPUTE AREA OF BOUNDING BOX
S44
MAIN SUBJECT COMPACTNESS FEATURE

FIG. 5D

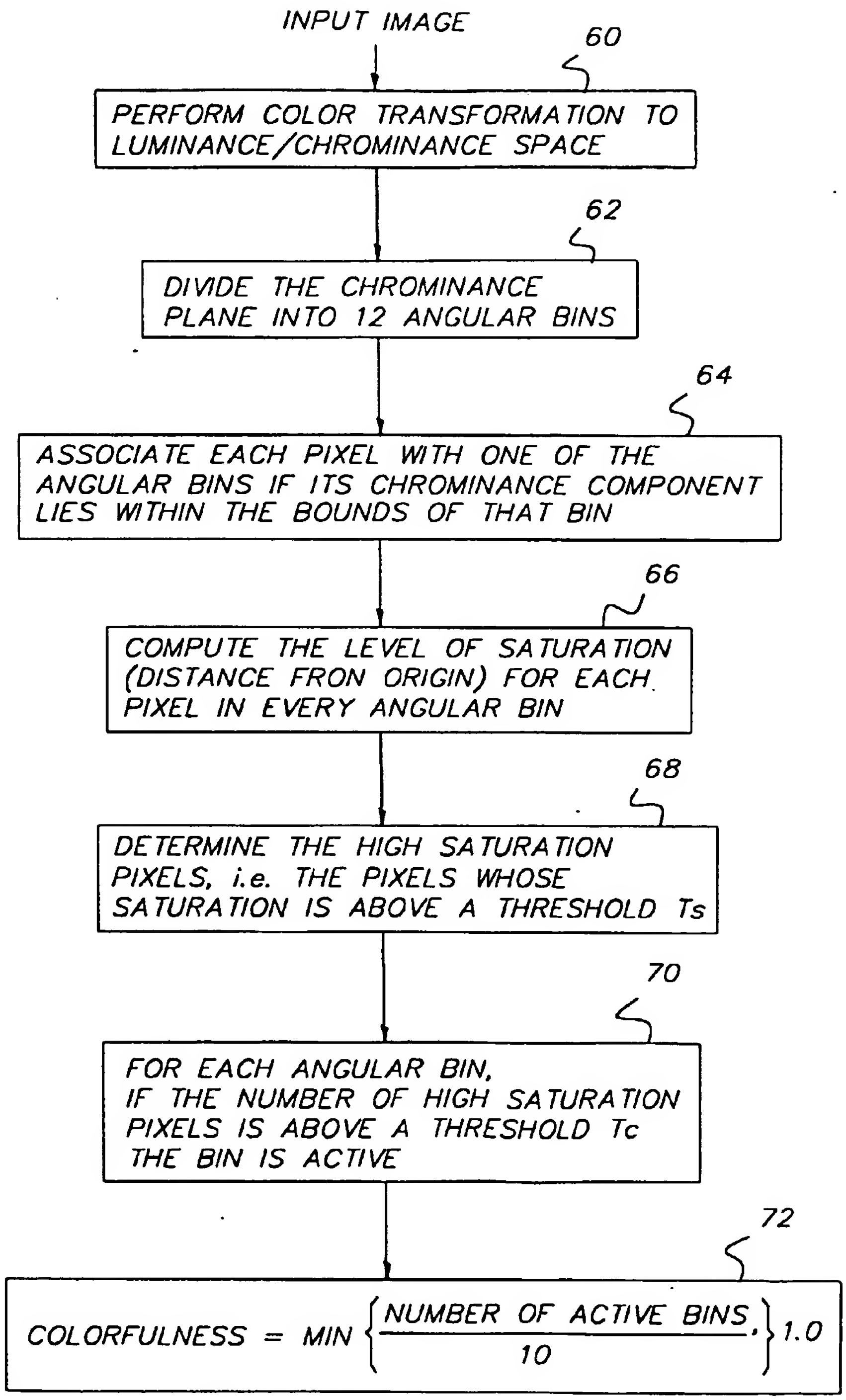
INPUT IMAGE
60
PERFORM COLOR TRANSFORMATION TO LUMINANCE/CHROMINANCE SPACE
62
DIVIDE THE CHROMINANCE PLANE INTO 12 ANGULAR BINS
64
ASSOCIATE EACH PIXEL WITH ONE OF THE ANGULAR BINS IF ITS CHROMINANCE COMPONENT LIES WITHIN THE BOUNDS OF THAT BIN
66
COMPUTE THE LEVEL OF SATURATION (DISTANCE FRON ORIGIN) FOR EACH PIXEL IN EVERY ANGULAR BIN
68
DETERMINE THE HIGH SATURATION PIXELS, i.e. THE PIXELS WHOSE SATURATION IS ABOVE A THRESHOLD Ts
70
FOR EACH ANGULAR BIN, IF THE NUMBER OF HIGH SATURATION PIXELS IS ABOVE A THRESHOLD Tc THE BIN IS ACTIVE
72
COLORFULNESS = MIN {NUMBER OF ACTIVE BINS / 10, 1.0}

FIG. 6

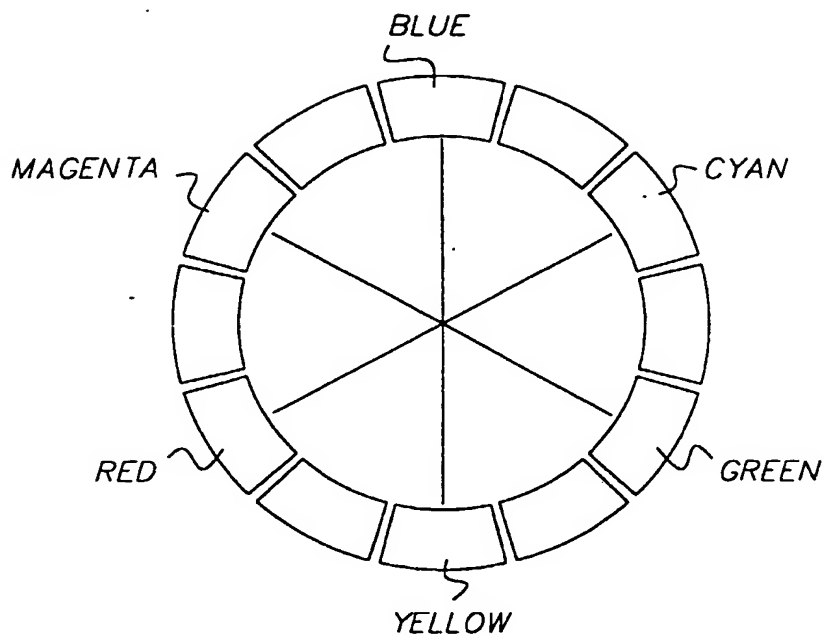
BLUE
MAGENTA
CYAN
RED
GREEN
YELLOW

FIG. 7

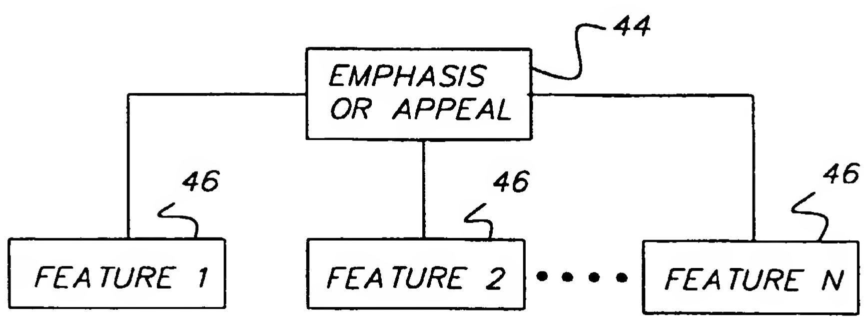
44
EMPHASIS OR APPEAL
46
46
46
FEATURE 1
FEATURE 2
FEATURE N

FIG. 10

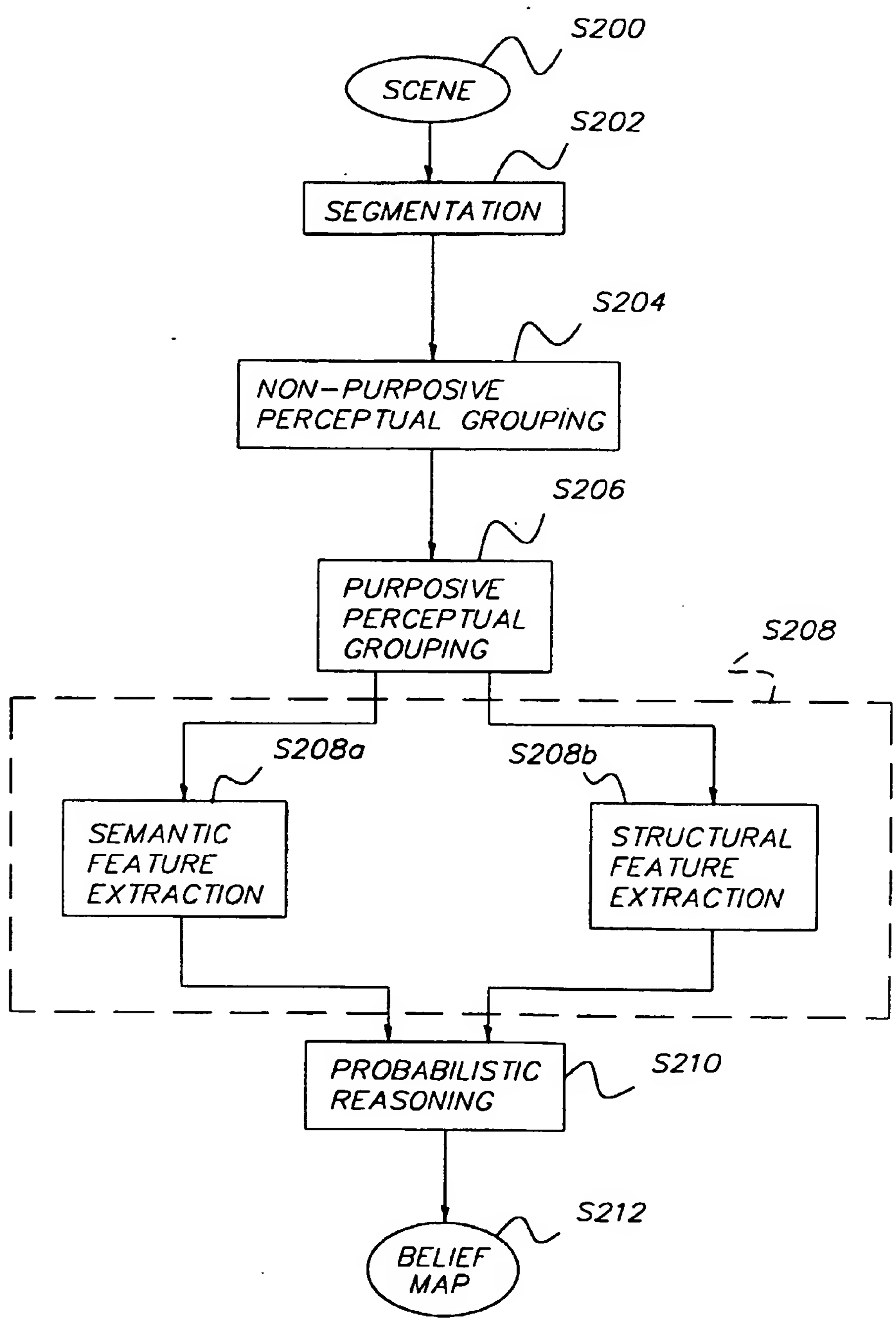
S200
SCENE
S202
SEGMENTATION
S204
NON-PURPOSIVE PERCEPTUAL GROUPING
S206
PURPOSIVE PERCEPTUAL GROUPING
S208
S208a
S208b
SEMANTIC FEATURE EXTRACTION
STRUCTURAL FEATURE EXTRACTION
PROBABILISTIC REASONING
S210
S212
BELIEF MAP

FIG. 9

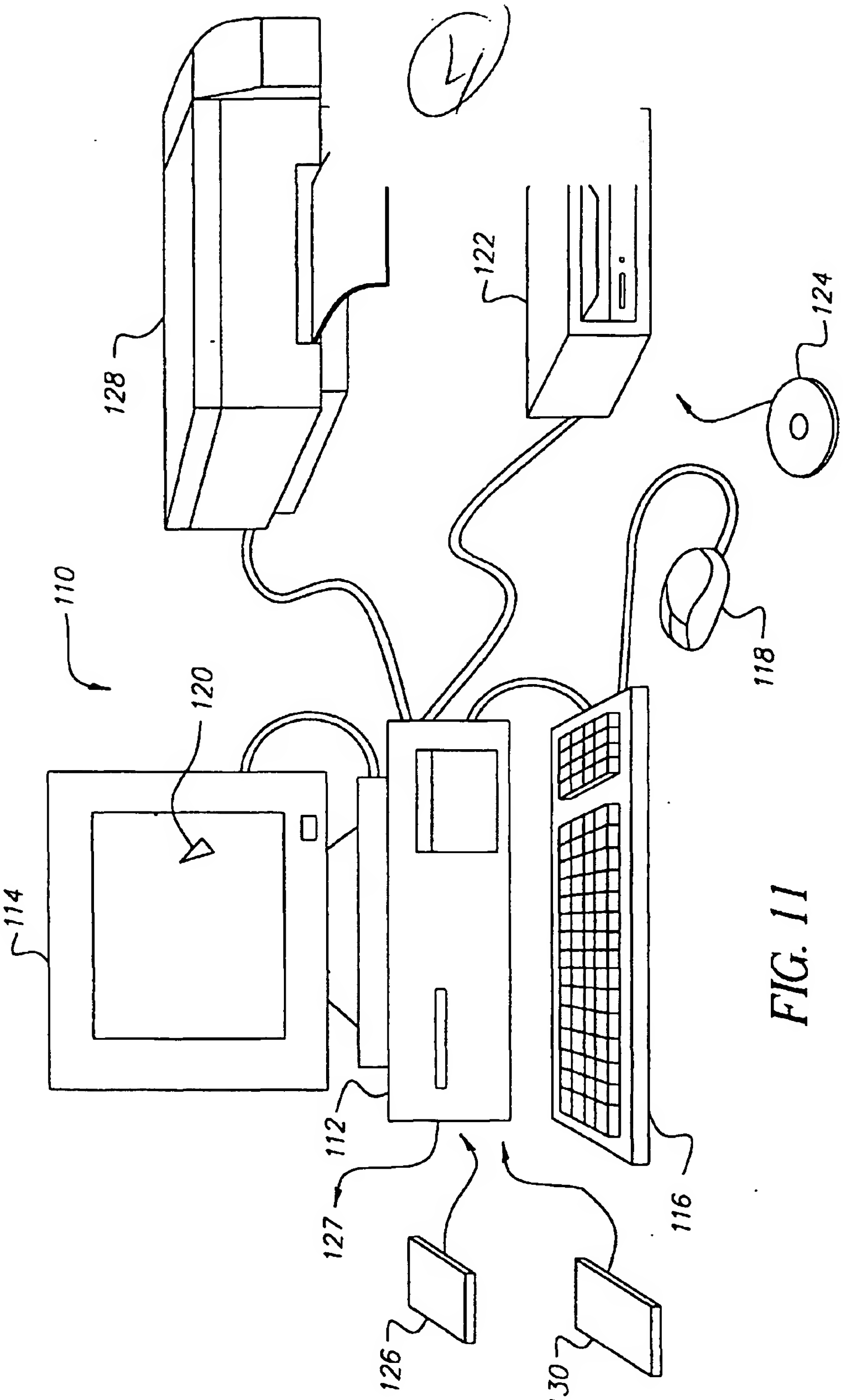


FIG. 11